

Genetically predicted serum vitamin D and COVID-19: a Mendelian randomisation study

Bonnie K Patchen ¹, Andrew G Clark,² Nathan Gaddis,³ Dana B Hancock,³ Patricia A Cassano ^{1,4}

To cite: Patchen BK, Clark AG, Gaddis N, *et al.* Genetically predicted serum vitamin D and COVID-19: a Mendelian randomisation study. *BMJ Nutrition, Prevention & Health* 2021;4:e000255. doi:10.1136/bmjnp-2021-000255

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjnp-2021-000255>).

¹Division of Nutritional Sciences, Cornell University, Ithaca, New York, USA

²Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York, USA

³GenOtics, Bioinformatics and Translational Research Center, Research Triangle Institute, Research Triangle Park, North Carolina, USA

⁴Population Health Sciences, Weill Cornell Medical College, New York, New York, USA

Correspondence to

Dr Patricia A Cassano, Division of Nutritional Sciences, Cornell University, Ithaca, NY 14853, USA; pac6@cornell.edu

Received 11 February 2021

Revised 9 April 2021

Accepted 10 April 2021

Published Online First

4 May 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

ABSTRACT

Objectives To investigate causality of the association of serum vitamin D with the risk and severity of COVID-19 infection.

Design Two-sample Mendelian randomisation study.

Setting Summary data from genome-wide analyses in the population-based UK Biobank and SUNLIGHT Consortium, applied to meta-analysed results of genome-wide analyses in the COVID-19 Host Genetics Initiative.

Participants 17 965 COVID-19 cases including 11 085 laboratory or physician-confirmed cases, 7885 hospitalised cases and 4336 severe respiratory cases, and 1 370 547 controls, primarily of European ancestry.

Exposures Genetically predicted variation in serum vitamin D status, instrumented by genome-wide significant single nucleotide polymorphisms (SNPs) associated with serum vitamin D or risk of vitamin D deficiency/insufficiency.

Main outcome measures Susceptibility to and severity of COVID-19 infection, including severe respiratory infection and hospitalisation.

Results Mendelian randomisation analysis, sufficiently powered to detect effects comparable to those seen in observational studies, provided little to no evidence for an effect of genetically predicted serum vitamin D on susceptibility to or severity of COVID-19 infection. Using SNPs in loci related to vitamin D metabolism as genetic instruments for serum vitamin D concentrations, the OR per SD higher serum vitamin D was 1.04 (95% CI 0.92 to 1.18) for any COVID-19 infection versus population controls, 1.05 (0.84 to 1.31) for hospitalised COVID-19 versus population controls, 0.96 (0.64 to 1.43) for severe respiratory COVID-19 versus population controls, 1.15 (0.99 to 1.35) for COVID-19 positive versus COVID-19 negative and 1.44 (0.75 to 2.78) for hospitalised COVID-19 versus non-hospitalised COVID-19. Results were similar in analyses using SNPs with genome-wide significant associations with serum vitamin D (ie, including SNPs in loci with no known relationship to vitamin D metabolism) and in analyses using SNPs with genome-wide significant associations with risk of vitamin D deficiency or insufficiency.

Conclusions These findings suggest that genetically predicted differences in long-term vitamin D nutritional status do not causally affect susceptibility to and severity of COVID-19 infection, and that associations observed in previous studies may have been driven by confounding. These results do not exclude the possibility of low-magnitude causal effects or causal effects of acute responses to therapeutic doses of vitamin D.

INTRODUCTION

The COVID-19 pandemic has reached every corner of the globe and continues to spread. With >133 million cases and nearly 2.9 million deaths globally at time of writing,¹ identification of risk factors for susceptibility to SARS-CoV-2 infection and severity of COVID-19 is critical. Vitamin D nutritional status is a promising modifiable risk factor, and higher vitamin D is posited to reduce the risk of SARS-CoV-2 infection and the severity of the clinical course of COVID-19. Hypothesised vitamin D effects are biologically plausible given prior evidence that vitamin D upregulates innate and adaptive immunity to fight infection and reduce inflammation,² is associated with a reduced risk of respiratory disease mortality³ and enhances expression of ACE2, which is hypothesised to modulate the immune system response to SARS-CoV-2 infection.^{4 5} A recent in vitro study showed that vitamin D reduces viral load of nasal epithelial cells infected with SARS-CoV-2,⁶ and preliminary evidence from two small human trials earlier in the pandemic suggested that vitamin D supplementation may help improve the prognosis of COVID-19 in infected individuals.^{7 8}

Ecological and observational studies lend further support to the hypothesis that lower serum vitamin D concentrations are associated with both an increased risk of COVID-19 and an increased severity of the infection. Early in the pandemic, the prevalence of COVID-19 in European countries was negatively correlated with national averages for serum vitamin D concentration.⁹ COVID-19 mortality rates were positively correlated with latitude, a proxy for UV B exposure (higher latitude associates with lower UV B), which is required for synthesis of vitamin D in the body.^{10 11} More recent studies report associations of lower prepandemic serum vitamin D concentration or higher risk of vitamin D

insufficiency with susceptibility to COVID-19 infection^{12–14} and with severity of COVID-19 infection in hospitalised patients as indicated by elevated biological markers of the cytokine storm, increased risk of intubation or use of supplemental oxygen, intensive care unit (ICU) admission and death.^{15–18} The mounting evidence from observational studies and the known effects of vitamin D on the immune system contribute to speculation about whether a simple and immediate intervention like vitamin D supplementation might be effective in reducing risk of COVID-19 infection or severity, but sources of guidance for clinicians and the public cite a lack of evidence for a causal association.^{19–20} Further evidence is urgently needed.

The uncertainty about the causal association of vitamin D and COVID-19 arises because of the well-known association of serum vitamin D concentration with many known risk factors for COVID-19, including age, sex, body mass index (BMI) and race/ethnicity. The largest observational study of the UK Biobank data (n=348 598 participants) reported little to no association of serum vitamin D with risk of COVID-19 infection in multivariate models adjusted for covariates.²¹ In the absence of large, rigorous randomised trials, it is challenging to determine whether other reported inverse associations of vitamin D and risk of COVID-19 are causal. Given the high global prevalence of vitamin D insufficiency, estimating the true effect of vitamin D on risk of COVID-19 infection and severity is important.

Mendelian randomisation (MR), a study design to address causality, uses a form of instrumental variable analysis to improve causal inference and to address the biases inherent in observational studies. MR uses genetic variants that are associated with the exposure (ie, serum vitamin D) as instrumental variables, or genetic instruments, which represent the long-term usual exposure. Following the laws of independent assortment, genetic variants are inherited from parent to offspring in a random and independent manner. An individual's genotype therefore mimics the lifelong randomisation of individuals into groups with different long-term serum vitamin D levels. Valid MR analysis depends on three key assumptions: (1) adequate strength and validity of the

genetic instruments (ie, genetic variants that underlie vitamin D metabolism), (2) independence of the genetic instruments from any confounders and (3) absence of direct effects of the genetic instruments on the outcome of interest (ie, COVID-19 status). MR addresses limitations in observational data including confounding, reverse causality and measurement error, and supports triangulation on the evidence for causality when randomised trials are either impossible to conduct or currently unavailable. Genome-wide association studies (GWAS) of serum vitamin D levels identified genetic variants that are robustly associated with serum vitamin D status in different populations and ancestries.^{22–28} Capitalising on this finding, researchers successfully applied MR to study associations of vitamin D with many clinical outcomes, including diseases of immune system dysregulation and inflammation.^{29–45} We used MR to investigate causality of the relationship of serum vitamin D status with the risk and severity of COVID-19 infection. One prior study used MR to study the vitamin D-COVID-19 association.⁴⁶ The present study adds new information by using multiple genetic instruments designed to maximise instrument strength and validity, validating the instruments across population subgroups, and considering additional COVID-19 clinical outcomes.

METHODS

Two-sample Mendelian randomisation with multiple instruments

We estimated the effect of serum vitamin D levels on risk of COVID-19 infection and severity with two-sample MR.⁴⁷ The associations of the genetic instruments with the exposure(s) and outcome(s) were estimated in separate, independent samples, then used to calculate the MR estimate of the effect of serum vitamin D status on COVID-19.⁴⁷ We used summary data from a GWAS of serum vitamin D in the UK Biobank (sample 1)²² and genome-wide data for patients with COVID-19 versus comparison groups from the COVID-19 Host Genetics Initiative (sample 2) (figure 1).⁴⁸ Because of overlap between the UK Biobank and the COVID-19 Host Genetics Initiative samples, which may bias MR results towards effects estimated from

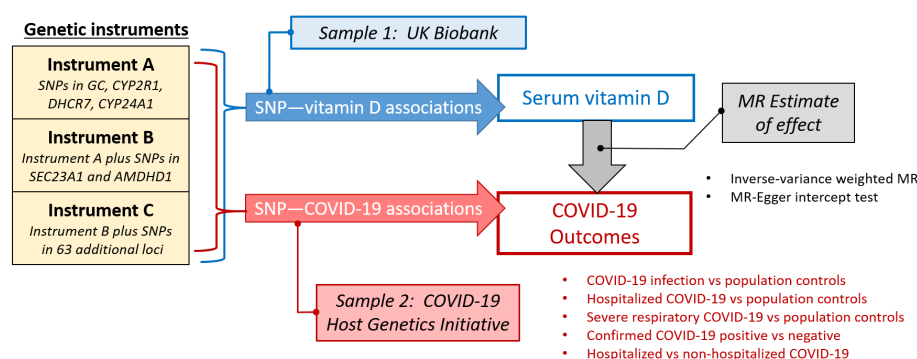


Figure 1 Schematic of two-sample Mendelian randomisation (MR) study design. SNP, single nucleotide polymorphism.

traditional observational studies,⁴⁹ we replicated the analyses using summary GWAS data for serum vitamin D from the SUNLIGHT Consortium.²³ We also performed sensitivity analyses to evaluate prior hypotheses about the direct effect of vitamin D deficiency or insufficiency on COVID-19 outcomes using summary data from a meta-analysis of associations of vitamin D SNPs with the dichotomous outcome of vitamin D deficiency versus sufficiency.²⁴

Vitamin D genetic instrument(s)

We constructed all genetic instruments for serum vitamin D levels based on independent genetic loci associated with serum vitamin D levels at genome-wide significance ($p < 5 \times 10^{-8}$). For the primary analysis, we used a biologically plausible genetic instrument (instrument A) consisting of SNPs in genetic loci that encode proteins in vitamin D transport and metabolism (online supplemental figure 1), including the vitamin D binding protein (*GC*), 25-OH hydroxylase (*CYP2R1*), 7-dehydrocholesterol reductase (*DHCR7*) and 24-hydroxylase (*CYP24A1*). All SNPs included in instrument A were associated with serum vitamin D at genome-wide significance in the UK Biobank and replicated in the SUNLIGHT Consortium.^{22 23}

In secondary analyses, we expanded the genetic instrument to include additional genome-wide significant SNPs in loci with no known relation to vitamin D metabolism (figure 1). The first expanded instrument (instrument B) included SNPs in two additional loci, *Sec23* homolog A (*SEC23A*) and amidohydrolase domain containing 1 (*AMDHD1*), that were associated with serum vitamin D at genome-wide significance in both the UK Biobank and the SUNLIGHT Consortium. The second expanded instrument (instrument C) included SNPs in 63 additional loci associated with serum vitamin D at genome-wide significance in the UK Biobank (online supplemental table 1).

Finally, we constructed separate genetic instruments for risk of vitamin D deficiency and insufficiency (defined as serum vitamin D < 50 nmol/L and < 75 nmol/L, respectively) based on meta-analysis results of vitamin D SNP-risk of vitamin D deficiency/insufficiency associations in the SUNLIGHT Consortium.²⁴ The genome-wide significant SNPs contributing to the vitamin D deficiency and insufficiency instruments were all in loci related to vitamin D metabolism, including *GC*, *DHCR7* and *CYP2R1*.

SNP-vitamin D associations

We extracted summary data, including beta-coefficients, SEs and effect alleles and their frequencies for each SNP contributing to the genetic instruments. We primarily used summary data from a UK Biobank GWAS of serum vitamin D in participants of 'white British' ancestry as defined by genotype principal component analysis ($n = 401\,460$)²² and replicated the analyses using summary data from a SUNLIGHT Consortium meta-analysis GWAS of serum vitamin D (31 European ancestry cohorts, $n = 79\,366$).²³ For sensitivity analyses, we used data from a SUNLIGHT Consortium meta-analysis of associations of

vitamin D SNPs with risk of vitamin D deficiency/insufficiency (four European ancestry cohorts, $n = 16\,905$).²⁴

Details of the UK Biobank and SUNLIGHT Consortia GWAS of serum vitamin D and the candidate SNP-vitamin D deficiency/insufficiency meta-analysis are described elsewhere.^{22–24} Briefly, in the UK Biobank GWAS, serum vitamin D data were log-transformed and standardised to a mean of 0 and SD of 1, and SNP-vitamin D association models were adjusted for age, sex, season of vitamin D measurement, vitamin D supplementation, genotype batch, genotype array and assessment centre (as a proxy for latitude). In the SUNLIGHT Consortium meta-analysis GWAS of serum vitamin D, serum vitamin D data were log-transformed and SNP-vitamin D association models were adjusted for age, sex, BMI, month of sample collection, cohort-specific variables such as geographical location and assay batch and genotype principal components.²² In the SUNLIGHT Consortium meta-analysis of SNP-risk of vitamin D deficiency/insufficiency associations, models were adjusted for age, sex, BMI and season, and genomic control was applied to control for population stratification.²⁴

Vitamin D instrument strength

We assessed the strength and validity of the genetic instruments for vitamin D by calculating the F-statistic,⁵⁰ according to the method described by Burgess *et al* for two-sample MR.⁴⁹ First, we calculated the variance in vitamin D explained by each SNP (R^2_{snp}) using the equation $R^2_{\text{snp}} = 2\alpha^2 \text{MAF}(1-\text{MAF})$, where α = SNP-serum vitamin D association and MAF = minor allele frequency. Next, we calculated the F-statistic for each of the instruments, using the equation $F = (n - K - 1) / K \times (R^2_{\text{instrument}}) / (1 - R^2_{\text{instrument}})$, where n = sample size, K = number of SNPs contributing to the genetic instrument and $R^2_{\text{instrument}} = \sum R^2_{\text{snp}}$ across SNPs contributing to the instrument. We based all calculations on GWAS data from the UK Biobank²² and calculated the F-statistic for sample sizes of 10 000 and 1.3 million individuals, the approximate minimum and maximum sample sizes for the COVID-19 outcomes. We considered the MR standard of F-statistic > 10 as an indicator of instrument strength.⁵⁰

Evaluation of potential pleiotropy of vitamin D instruments

We used LDlink⁵¹ to evaluate all SNPs included in instruments A, B and C, as well as SNPs in linkage disequilibrium (LD) with those SNPs ($R^2 \geq 0.8$ based on 1000 Genomes CEU and GBR populations), for genome-wide significant associations with phenotypes other than vitamin D. We further explored potential pleiotropy with COVID-19 risk factors by extracting summary statistics for associations of SNPs in the six vitamin D-associated loci replicated across the UKBB and SUNLIGHT Consortium (instrument B) with BMI and smoking phenotypes, including BMI, waist circumference, obesity, smoking initiation, smoking cessation and cigarettes smoked per day. BMI summary statistics were from a GWAS and meta-analysis in the Genetic Epidemiology Research on Adult Health

and Ageing (GERA) cohort and Genetic Investigation of Anthropomorphic Traits (GIANT) consortium combined with the UKBB and were accessed through the MR-base platform (study ID ebi-a-GCST006368).⁵² Waist circumference summary statistics were from a GWAS meta-analysis of 57 independent European ancestry cohorts and were accessed through the MR-base platform (study ID ieu-a-60).⁵³ Obesity summary statistics were from the FinnGen Biobank and were accessed through the MR-base platform (study ID finn-a-E4_OBESITY). Summary statistics for smoking initiation, smoking cessation and cigarettes smoked per day were from a meta-analysis of 30 European ancestry studies in the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN) consortium and were accessed through the University of Minnesota Data Repository (<https://doi.org/10.13020/3b1n-ff32>).⁵⁴ The threshold for declaring statistical significance of SNP associations with BMI and smoking phenotypes was $p < 0.0083$ (0.05/6 vitamin D SNPs) to account for multiple testing using the Bonferroni correction.

Genetic instruments for multivariable MR conditioning on BMI

Based on the results of the targeted look up of vitamin D-associated SNPs with BMI and smoking phenotypes (online supplemental table 2), we performed summary statistic multivariable MR (MVMR) analysis to estimate the effect of vitamin D on COVID-19 outcomes conditioning on BMI. Genetic instruments for BMI were identified through the MR-base platform as SNPs associated with BMI at genome-wide significance in the GERA/GIANT and UKBB European ancestry GWAS and meta-analysis (study ID ebi-a-GCST006368) and a mixed ancestry meta-analysis of 82 GWAS and Metabochip analyses (study ID ieu-a-2).⁵⁵ MVMR analysis estimating the direct effect of vitamin D on COVID-19 outcomes conditioning on BMI was performed separately using vitamin D genetic instruments from the UKBB and SUNLIGHT Consortium and BMI genetic instruments from the European and mixed ancestry BMI GWAS.

SNP-COVID-19 associations

We used genome-wide data for COVID-19 cases and comparison groups from the COVID-19 Host Genetics Initiative, which is described elsewhere.⁴⁸ Briefly, the initiative comprises a global effort to study genetic contributions to variability in COVID-19 infection and severity. Contributing studies performed genome-wide analyses for COVID-19 case and comparison groups following an analysis plan developed by the initiative. Studies with multiple ancestries performed stratified analyses, and all analyses were adjusted for sex, age, age², sex \times age interaction and genotype principal components. Results were summarised across studies and ancestries by meta-analysis.

The publicly available summary results were downloaded from the COVID-19 Host Genetics Initiative's fourth data freeze and meta-analysis (<https://www.covid19hg.org/results/>), which were released on 20 October 2020. We considered the following five

COVID-19 case versus control comparisons, using definitions provided by the initiative: (1) any COVID-19 versus population controls, (2) hospitalised COVID-19 versus population controls, (3) very severe respiratory confirmed COVID-19 versus population controls, (4) any COVID-19 versus confirmed COVID-19 negative and (5) hospitalised versus non-hospitalised COVID-19. For each COVID-19 'case' versus comparison group, we extracted summary data, including beta-coefficients, SEs, effect alleles and effect allele frequencies corresponding to the SNPs included in the vitamin D genetic instruments. For SNPs that were not available in the public data, we identified proxy SNPs based on LD, using the standard MR threshold for LD proxies of $r^2 > 0.80$.⁵⁶ All LD proxy SNPs were defined using 1000 Genomes European (CEU and GBR) sample data and identified through LDlink.⁵¹ Case and comparison group definitions, sample sizes and ancestry distributions for the different comparisons are summarised in table 1. Studies contributing to these analyses are listed in online supplemental table 3.

Summary data were available for mixed ancestry analyses for all outcomes. For the outcomes of COVID-19 versus population controls and hospitalised COVID-19 versus population controls, summary data from analyses limited to European ancestry participants were also available. For these outcomes, we extracted summary data from both the mixed ancestry and the European ancestry analyses for comparison and to address differences in ancestry composition between vitamin D GWAS and COVID-19 Host Genetics Initiative sample populations.

Stratification of SNP-vitamin D associations by COVID-19 risk factors

Our study design assumes that the potential effect of vitamin D nutritional status on susceptibility to and severity of COVID-19 infection is the same across different COVID-19 risk categories, including ancestry, age, sex, BMI and smoking status. MR analysis under this assumption requires the genetic instruments for vitamin D to be robust across the COVID-19 risk subgroups. To address this assumption, we conducted stratified analyses of the SNP-vitamin D associations for all SNPs contributing to the genetic instruments using data from the UK Biobank. UK Biobank Resource data—including serum vitamin D levels, vitamin D supplementation, covariates and stratification variables and imputed genotype dosages with reference to Haplotype Reference Consortium and UK10K haplotype resource—were obtained under Application Number 24603.

The SNP-vitamin D associations were estimated stratified by sex (male vs female), age group (40–50 vs 50–60 vs 60–70), BMI category (18.5–25 vs 25–30 vs 30–40 vs >40) and smoking status (never vs former vs current smokers). Analyses were performed using the GENESIS R/Bioconductor package,⁵⁷ separately for European ancestry ($n=421\,407$) and African ancestry ($n=78\,59$) participants, as classified via analysis with STRUCTION⁵⁸ in comparison with the CEU, YRI and CHB 1000

Table 1 Sample description for COVID-19 case and comparison groups

Comparison	Case definition	Comparison group definition	Sample size	Ancestry
Any COVID-19 case versus population controls	Laboratory/Physician confirmed OR self-reported positive	Everyone that is not a case (eg, the population)	17 965 cases, 1370547 controls	93.6% EUR 3.1% EAS/SAS 1.8% AFR 1% ARAB <1% AMR <1% HIS
Hospitalised COVID-19 cases versus population controls	Laboratory/Physician confirmed AND hospitalised for COVID-19	Everyone that is not a case (eg, the population)	7885 cases, 961 804 controls	93.7% EUR 3.5% EAS/SAS 1.1% AFR 1.4% ARAB <1% AMR <1% HIS
Severe respiratory COVID-19 cases versus population controls	Laboratory confirmed AND hospitalised AND respiratory support or death	Everyone that is not a case (eg, the population)	4336 cases, 623 902 controls	99.6% EUR <1% AMR
COVID-19-positive cases versus COVID-19-negative controls	Laboratory/Physician confirmed OR self-reported positive	Laboratory confirmed OR self-reported negative	11 085 cases, 116 794 controls	86.5% EUR <1% EAS/SAS 9.6% AFR 3.2% HIS
Hospitalised versus non-hospitalised COVID-19	Laboratory/Physician confirmed AND hospitalised for COVID-19	Laboratory or physician confirmed COVID-19	2430 hospitalised 8478 non-hospitalised	75.3% EUR <1% EAS/SAS 12.6% AFR 6.4% ARAB 4.7% HIS

AFR, African; AMR, Ad Mixed American; ARAB, Arabic; EAS/SAS, East Asian/South Asian; EUR, European; HIS, Hispanic.

Genomes populations.^{59 60} Statistical models resembled those described for SNP-vitamin D associations in UK Biobank,²² except that genotype batch was not included as a covariate and vitamin D levels were not log-transformed.

Statistical analysis for two-sample MR

We performed two-sample MR using the inverse variance weighted (IVW) method, which assumes that pleiotropy is either non-existent or balanced (ie, any associations of genetic instrument SNPs with phenotypes other than vitamin D are randomly positive and negative such that the mean pleiotropic effect is zero).⁶¹ We used the MR-Egger intercept test to check for evidence of directional pleiotropy, and performed sensitivity analyses using the MR-Egger, weighted-median and weighted-mode methods.⁶²⁻⁶⁴ The MR-Egger method relaxes the assumption of no directional pleiotropy and corrects for pleiotropy-induced bias by allowing for a non-zero intercept.^{61 62} The weighted-median method assumes that at least 50% of the weight of the genetic instrument comes from non-pleiotropic SNPs and produces an unbiased estimate so long as <50% of SNPs have pleiotropic effects.⁶³ The weighted-mode method assumes that the most common (modal) SNP-vitamin D effect is the true effect and therefore allows for the inclusion of pleiotropic SNPs without biasing the MR estimate.⁶⁴ Where evidence of directional pleiotropy existed (MR-Egger intercept test $p < 0.05$), we prioritised the MR estimates from the MR-Egger, mode-weighted and median-weighted

analyses, otherwise we prioritised the estimates from the IVW method. We further explored potential pleiotropy of vitamin D instruments with additional sensitivity analyses, including IVW MR of vitamin D and COVID-19 outcomes with a genetic instrument limited to SNPs with no known genome-wide significant associations other than with vitamin D, and MVMR of vitamin D and COVID-19 outcomes conditioning on genetically predicted BMI. Results for our primary analysis are presented as ORs and 95% CIs for COVID-19 case versus comparator per SD increase in log-transformed serum vitamin D. The threshold for statistical significance was set at a p value < 0.05 . All analyses were performed using R Studio V.1.2.1335 and the 'TwoSampleMR' R package (<https://github.com/MRCIEU/TwoSampleMR>).⁵⁶

We estimated potential bias under the null due to sample overlap between the UK Biobank vitamin D GWAS and the COVID-19 Host Genetics Initiative using estimates of the vitamin D-COVID-19 associations from observational studies, the percentage of sample overlap and the relative bias (reciprocal of the F-statistic), as previously described.⁴⁹ We compared SNP-vitamin D associations stratified by sex, age group, BMI category and smoking status with Pearson's correlation coefficients.

We calculated the minimum OR detectable at 80% power for each COVID-19 outcome using the 'mRnd' tool (<https://shiny.cnsgenomics.com/mRnd/>).⁶⁵ Based on heritability estimates from the UK Biobank

Table 2 Characteristics of vitamin D instruments and association with measured vitamin D status in UK Biobank (n=401 460)

Loci	SNP	EA	NEA	EAF	Beta-vitD	P-vitD	F-statistic	
							N=10 000	N=1 300 000
Instrument A							53	6915
GC	rs11723621	G	A	0.29	−0.19	2.9E-1689	147	19072
CYP2R1	rs10832289	T	A	0.41	−0.07	2.03E-266	23	2999
DHCR7	rs12803256	G	A	0.78	0.10	1.3E-378	35	4597
CYP24A1	rs6127099	T	A	0.28	−0.04	9.30E-62	5	715
Instrument B							37	4746
SEC23A	rs8018720	C	G	0.82	−0.03	4.04E-36	3	391
AMDHD1	rs10859995	C	T	0.58	−0.04	7.03E-89	8	983
Instrument C*							5	600

Source: Manousaki *et al.*²²

*Characteristics of all SNPs in instrument C are shown in online supplemental table 1.

Beta-vitD, beta-coefficient for association of an additional effect allele with an SD change in serum vitamin D in the log scale; EA, effect allele; EAF, effect allele frequency; F-statistic, measure of instrument strength, calculated as described in the 'Methods' section, shown for approximate minimum and maximum sample sizes across the COVID-19 outcomes; NEA, non-effect allele; P-vitD, p value for the beta-coefficient; SNP, single nucleotide polymorphism.

and SUNLIGHT Consortium vitamin D GWAS,^{22 23} we assumed the genetic instruments accounted for 3% of the variance in serum vitamin D for all power calculations.

Patient and public involvement

We used publicly available data from several consortia including the UK Biobank, the SUNLIGHT Consortium and the COVID-19 Host Genetics Initiative in this study. These consortia were not involved in any stage of the design and conduct of this research, nor were they asked to advise on interpretation or writing up of results. No patients were involved in any stage of the research process. There are no plans to disseminate the results of the research to study participants or the relevant patient community.

RESULTS

Genetic instruments for vitamin D

Characteristics of the genetic instruments for serum vitamin D, including contributing SNPs, SNP-vitamin D associations in the UK Biobank and F-statistics, indicate that instruments A and B clearly exceed the MR standard for instrument strength of F-statistic >10 across the range of COVID-19 outcome sample sizes (table 2, online supplemental table 1). The F-statistics for instruments A, B and C were respectively 53, 37 and 5 for the minimum sample size of 10 000 participants and 6915, 4746 and 600 for the maximum sample size of 1.3 million participants. Instrument A, which was limited to SNPs in loci related to vitamin D metabolism (online supplemental figure 1), had the largest F-statistics across all sample sizes, reflecting the strength of the SNP-vitamin D associations for the SNPs in loci related to the vitamin D metabolic pathway.

COVID-19 case-comparison groups and power calculations

We considered multiple case definitions, sample sizes and ancestral compositions for the COVID-19 case-control comparisons in our analysis (table 1). Sample sizes across the analyses ranged from just over 10 000 (for hospitalised COVID-19 vs non-hospitalised COVID-19) to over 1.3 million (for COVID-19 vs population). The majority (>75%) of participants were of European ancestry. Non-European ancestry participants made up a much smaller fraction of the total sample.

Based on the sample sizes across the COVID-19 case-control comparisons, and assuming our genetic instruments explained 3% of the variance in serum vitamin D, we calculated that this study had 80% power to detect moderate effect sizes comparable to those seen in observational studies. The minimum detectable ORs for an SD increase in serum vitamin D ranged from 0.66 (hospitalised COVID-19 vs non-hospitalised COVID-19) to 0.88 (COVID-19 vs population; online supplemental table 4).

MR estimates for serum vitamin D effect on COVID-19 outcomes

MR estimates (ORs and 95% CIs) for the effect of genetically predicted serum vitamin D on risk of COVID-19 outcomes, calculated separately for instruments A, B and C, are shown in figure 2 and table 3. MR scatter plots and forest plots further illustrating these results are shown in online supplemental figure 2. MR-Egger intercept tests for pleiotropy were not statistically significant (MR-Egger intercept p>0.1) for all instruments and all outcomes (table 3). Thus, we prioritised the MR estimates from the IVW analyses, and found little to no evidence for an effect of vitamin D on risk of any COVID-19 outcome considered (p>0.1 for all comparisons). Results were similar for instruments B and C. Results of MR analyses limited to European ancestry participants, which were performed for the outcomes of COVID-19 versus population and

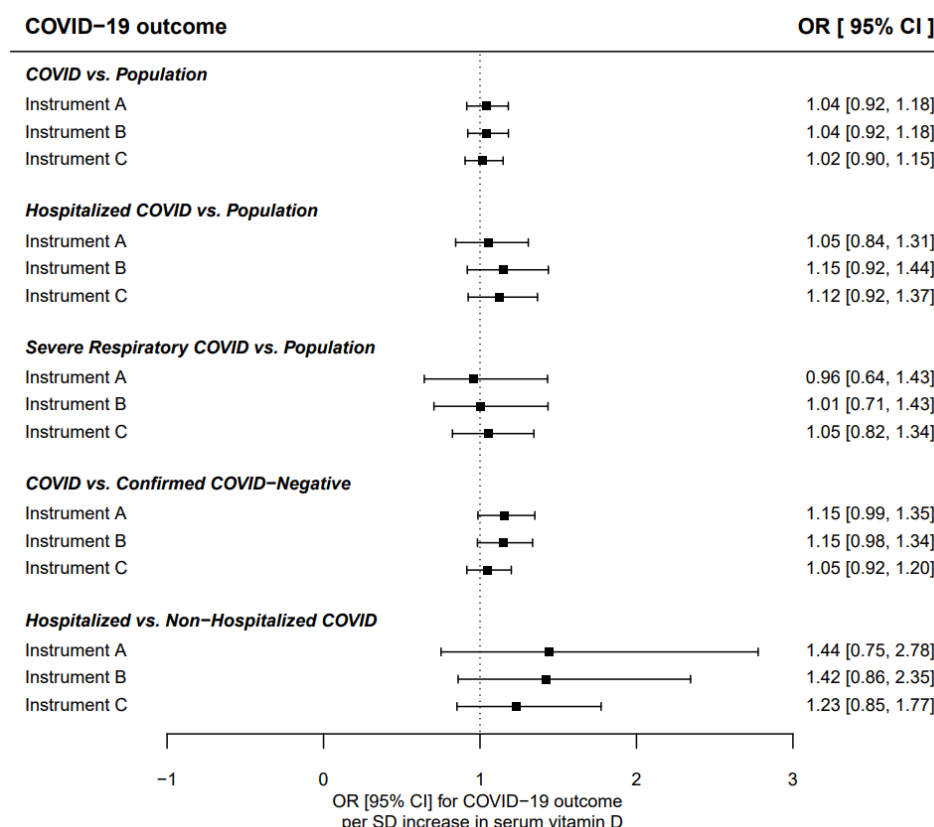


Figure 2 Inverse variance weighted Mendelian randomisation estimates for effect of serum vitamin D on risk of COVID-19 outcomes.

hospitalised COVID-19 versus population, were also non-significant (online supplemental figure 3).

For all outcomes, the MR-Egger intercept test showed little to no evidence of pleiotropy (table 3). Nevertheless, we performed several sensitivity analyses to explore potential pleiotropy. These included using the MR-Egger, weighted-median and weighted-mode MR methods, performing MVMR analysis of vitamin D conditioning on BMI, a known risk factor for COVID-19, and performing IVW MR with a genetic instrument limited to SNPs with no known genome-wide significant associations other than with vitamin D phenotypes. These analyses suggested a *direct* association of serum vitamin D with risk of hospitalised COVID-19 versus non-hospitalised COVID-19 (online supplemental figure 4, online supplemental table 5). For this outcome, MR estimates from the MR-Egger and mode-weighted analyses found that chronically *higher* vitamin D levels increase the odds of hospitalisation in COVID-19-infected participants (OR (95% CI) for instrument A: 1.83 (1.12 to 2.97) for mode-weighted MR and 3.69 (1.53 to 8.93) for MR-Egger). Estimates from MVMR analysis conditioning on BMI using data from three different BMI GWAS were similar in direction and magnitude (p ranging 0.05–0.14 for the three GWAS). MR estimates from the MR-Egger, mode-weighted, median-weighted and MVMR MR analyses were not statistically significant for all other outcomes, consistent with results from the IVW analysis. Results from MR analysis limiting

the genetic instrument to SNPs with no known genome-wide significant associations other than with vitamin D were consistent with results from the primary analyses (online supplemental table 6). Instrument strength for this analysis remained high for outcomes with higher numbers, but results for the outcomes with the lowest case prevalence (severe respiratory COVID-19 vs the population) and smallest sample size (hospitalised vs non-hospitalised COVID-19) may have been affected by weak instrument bias (instrument F-statistic <10) and/or underpowered (<80% power to detect effects comparable to those seen in observational studies).

Estimations of bias because of participant overlap between the UK Biobank and the COVID-19 Host Genetics Initiative were negligible. For the outcome with maximum participant overlap (58% for severe respiratory COVID-19 vs population), we estimated the bias for instruments A, B and C to be 0.1%, 0.2% and 1.0%, respectively (data not shown). We confirmed the negligible effect of participant overlap in our study by repeating the analyses using summary data for GWAS significant SNP-vitamin D associations from the SUNLIGHT Consortium in place of UK Biobank data (online supplemental table 7). Results using the SUNLIGHT Consortium data were consistent with results using the UK Biobank data; there was no evidence of pleiotropy for any outcome (MR-Egger intercept p>0.1), and IVW MR estimates for the effect of genetically predicted serum vitamin D on risk of COVID-19

Table 3 Mendelian randomisation (MR) estimates of effect of vitamin D on COVID-19 outcomes

MR method	COVID versus population			Hospitalised COVID versus population			Severe respiratory COVID versus population			COVID positive versus COVID negative			Hospitalised COVID versus non-hospitalised COVID		
	OR	95% CI		OR	95% CI		OR	95% CI		OR	95% CI		OR	95% CI	
Instrument A															
IVW	1.04	0.92 to 1.18		1.05	0.84 to 1.31		0.96	0.64 to 1.43		1.15	0.99 to 1.35		1.44	0.75 to 2.78	
MR-Egger intercept test	P=0.57			P=0.19			P=0.14			P=0.88			P=0.14		
Instrument B															
IVW	1.04	0.92 to 1.18		1.15	0.92 to 1.44		1.01	0.71 to 1.43		1.15	0.98 to 1.34		1.42	0.86 to 2.35	
MR-Egger intercept test	P=0.62			P=0.15			P=0.06			P=0.61			P=0.14		
Instrument C															
IVW	1.02	0.90 to 1.15		1.12	0.92 to 1.37		1.05	0.82 to 1.34		1.05	0.92 to 1.20		1.23	0.85 to 1.77	
MR-Egger intercept test	P=0.88			P=0.85			P=0.23			P=0.12			P=0.14		
Instrument of risk of vitamin D deficiency															
IVW	1.00	0.99 to 1.01		1.00	0.98 to 1.02		1.02	0.99 to 1.05		0.99	0.97 to 1.00		0.97	0.90 to 1.05	
MR-Egger intercept test	P=0.95			P=0.61			P=0.92			P=0.46			P=0.21		
Instrument of risk of vitamin D insufficiency															
IVW	1.00	0.99 to 1.01		1.00	0.98 to 1.02		1.06	1.00 to 1.05		0.99	0.98 to 1.00		0.97	0.90 to 1.05	
MR-Egger intercept test	P=0.87			P=0.64			P=0.76			P=0.59			P=0.22		

OR representing change in risk of outcome per SD increase in log-transformed serum vitamin D (instruments A, B and C) or difference in risk of outcome for individuals at risk of vitamin D deficiency/insufficiency versus individuals with sufficient vitamin D.
IVW, inverse variance weighted.

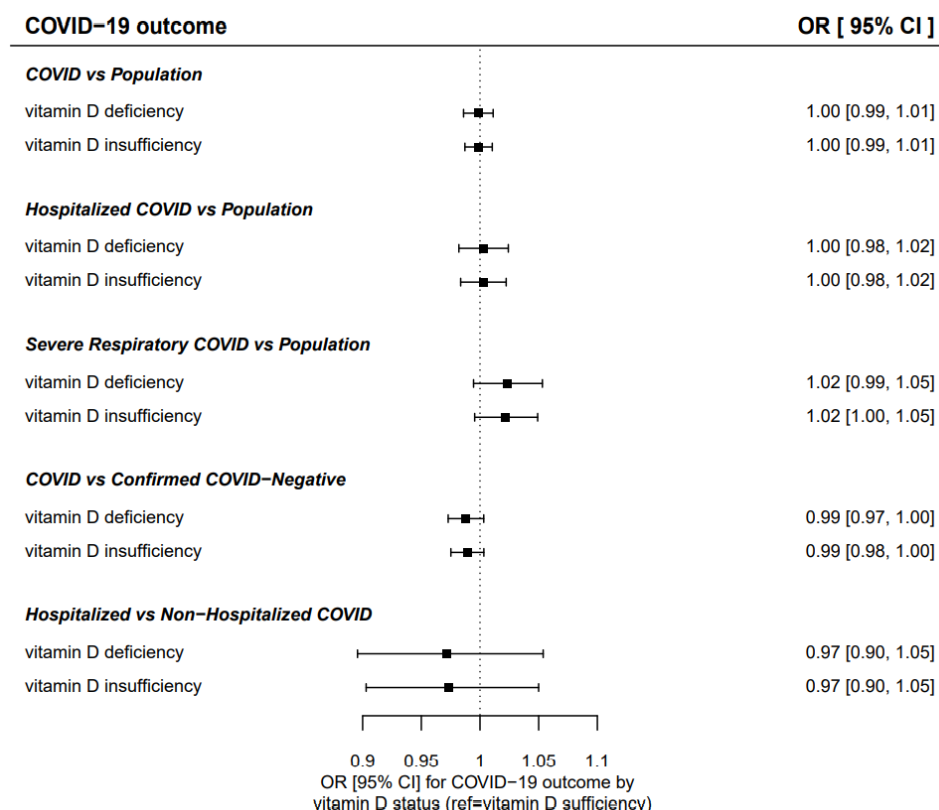


Figure 3 Inverse variance weighted Mendelian randomisation estimates for effect of vitamin D insufficiency or deficiency on risk of COVID-19 outcomes.

outcomes were similar in magnitude and direction and were not statistically significant.

MR estimates for vitamin D deficiency/insufficiency and COVID-19 outcomes

A further sensitivity analysis evaluated the effect of genetically predicted vitamin D deficiency and insufficiency on COVID-19 outcomes (table 3, figure 3). MR-Egger intercept tests for pleiotropy were all non-significant ($p>0.1$), and the IVW MR estimates did not support an effect of long-term vitamin D deficiency or insufficiency on the risk of any of the COVID-19 outcomes evaluated.

Covariate effects on SNP-vitamin D associations

Stratified analyses of SNP-vitamin D associations evaluated the validity of the genetic instruments across known COVID-19 risk factors, including ancestry, age, sex, BMI and smoking status. Correlations of age-specific, sex-specific, BMI-specific and smoking status-specific SNP-vitamin D associations are shown separately for European ($n=421\,407$) and African ($n=7859$) ancestry participants in online supplemental figure 5. For European ancestry participants, subgroup-specific SNP-vitamin D associations were strongly correlated across age, sex, BMI category and smoking status (Pearson's correlation coefficients ranging from 0.90 to 0.99), supporting the use of the instrument in the general European ancestry population. For African ancestry participants, the correlations among subgroup-specific SNP-vitamin D associations

were less clear (Pearson's correlation coefficients ranged from -0.95 to 0.72), suggesting that optimisation of the genetic instruments for the African ancestry population may be necessary. Given the small sample sizes in the African ancestry subgroups, further research is needed to validate the instruments and their use in African and other non-European ancestry populations.

DISCUSSION

We used summary data from genome-wide analysis in the UK Biobank, SUNLIGHT Consortium and COVID-19 Host Genetics Initiative to conduct a comprehensive MR study of serum vitamin D and COVID-19 outcomes. Although our study was powered to detect effects comparable to those seen in observational studies of vitamin D and COVID-19, we found little to no evidence of an effect of genetically predicted serum vitamin D levels on COVID-19 outcomes, including risk of COVID-19 infection, severe respiratory infection and hospitalisation. Our results were robust to multiple genetic instruments for vitamin D and were replicated using SNP-vitamin D association data from two independent samples. Results from sensitivity analyses evaluating the effect of genetically predicted vitamin D deficiency or insufficiency on risk of COVID-19 outcomes were similarly null. Our study builds on the findings of a prior MR study of vitamin D and COVID-19,⁴⁶ providing replication with genetic

instruments specifically designed to maximise instrument strength and address MR assumptions, and expanding the scope to include additional COVID-19 outcomes and individuals of non-European ancestry. In summary, our results suggest that long-term usual vitamin D nutritional status does not have a causal effect on susceptibility to COVID-19 infection and its severity.

Our results are consistent with findings from the largest observational study to date of serum vitamin D and COVID-19 infection,²¹ which used UK Biobank data on approximately 350 000 white, black and South Asian participants, and found associations of prepandemic vitamin D levels with COVID-19 infection in univariate models, but not in models adjusting for confounding by known COVID-19 risk factors, including age, sex, race/ethnicity, BMI, socioeconomic status, smoking status, diabetes, blood pressure, chronic illness and disability. We used the MR approach as an alternative way to address potential confounding. We instrumented long-term serum vitamin D status with genetic variants that are unlikely to have direct associations with known risk factors for COVID-19 infection. Specifically, the genetic instrument for our primary analysis (instrument A) was limited to SNPs in loci with functional links to serum vitamin D, including *GC*, *DHCR7*, *CYP2R1* and *CYP24A1*. With this approach, we saw no evidence for an effect of genetically predicted (ie, unconfounded) variation in serum vitamin D on risk and severity of COVID-19 infection.

These results are contrary to findings from several smaller observational studies showing an association of lower vitamin D levels with higher risk of testing positive for COVID-19 in the general population^{12–14} and higher risk of COVID-19-related hospitalisation, ICU admission, intubation and death among infected individuals.^{15–18} However, these studies were inconsistent in the covariates included in statistical models, and in several studies estimates of vitamin D associations with susceptibility to and severity of COVID-19 infection were attenuated with covariate adjustment, suggesting the observed effects could be the result of residual confounding. Additionally, many of these studies were case-control or cross-sectional studies, making it difficult to rule out the possibility of reverse causality (ie, that COVID-19 infection or its symptoms could lead to lowering of serum vitamin D). The MR approach addresses key limitations of these studies, including confounding and reverse causality, and is recognised as a method for improving causal inference in epidemiology.

This study has many strengths. We used summary data from the largest GWA studies of serum vitamin D to maximise the strength and validity of the genetic instruments for vitamin D and leveraged publicly available data on nearly 1.4 million participants, including 17 965 COVID-19 cases, from 38 study cohorts in the COVID-19 Host Genetics Initiative. We addressed potential pleiotropy of the genetic instruments, which if present would limit the interpretation of the MR results, by comparing results across the biologically plausible and expanded

instruments, testing for evidence of pleiotropy with the MR-Egger intercept test, and performing sensitivity analysis with alternative MR methods that address pleiotropy. We explored potential threshold effects using genetic instruments for vitamin D deficiency and insufficiency. Our findings were robust across all exploration, and, in triangulation with results of the largest to date observational study of vitamin D and COVID-19, the findings help to clarify the nature of the association of vitamin D with COVID-19 outcomes.

As in any MR study, the valid interpretation of our results rests on the MR assumptions of adequate strength of the genetic instruments for vitamin D and absence of direct effects of the instruments on COVID-19 outcomes and potential confounders (ie, pleiotropy). Our instruments easily met the MR standard for instrument strength (F-statistic >10) and we found no evidence of pleiotropy from the MR-Egger intercept test. Our results were consistent across the biologically plausible and expanded genetic instruments and largely similar in sensitivity analysis using MR methods designed to address pleiotropy. While the findings in our study are unlikely to be biased by pleiotropy, exhaustive tests for pleiotropy are challenging, making the MR assumptions difficult to fully verify.

We used a two-sample MR approach, using estimates of associations of the genetic instruments with vitamin D and with COVID-19 outcomes from two separate samples. This approach maximises sample size and reduces the likelihood of bias towards estimates from observational studies,⁴⁷ but requires that the two samples are drawn from similar populations. In our study, the summary data for associations of the genetic instruments with vitamin D were from analyses in European ancestry participants, while the data for associations of the genetic instruments with COVID-19 outcomes were from analyses including up to 25% non-European ancestry participants. Repeating the analysis with a subgroup of European-only participants for two outcomes, COVID-19 versus population controls and hospitalised COVID-19 versus population controls, indicated minimal impact of ancestral differences on MR estimates. However, ancestry-specific data were not available for COVID-19 outcomes with larger proportions of non-European ancestry participants, and the impact of ancestral differences could not be fully explored. In sensitivity analyses using alternative MR methods designed to address known and unknown pleiotropy, findings indicated an association of *higher* vitamin D levels with *higher* risk of hospitalisation among COVID-19-infected individuals, the outcome with the highest proportion of non-European ancestry participants. This finding could be driven by racial/ethnic disparities in access to healthcare (ie, hospitalised individuals with COVID-19 may be more likely of European ancestry and therefore have higher vitamin D status). We also found evidence for racial differences in risk-factor stratified SNP-vitamin D associations. Our results suggest that SNP-vitamin D associations are strongly correlated across sex, age, BMI and smoking subgroups for individuals of European ancestry, but not

for individuals of African ancestry, although because of small sample sizes for the African ancestry subgroups we are unable to draw firm conclusions. Further research using data from ancestry-specific GWAS of vitamin D and COVID-19 outcomes is warranted to further investigate the impact of race/ethnicity on the relation of vitamin D to COVID-19 risk.

Our MR study investigated causal inference questions raised by published observational studies of the relation of the total serum vitamin D biomarker to COVID-19 outcomes. Total serum vitamin D includes vitamin D bound to the vitamin D binding protein (85%), vitamin D bound to albumin (15%) and free vitamin D (<1%).⁶⁶ Recent studies suggest that free or active serum vitamin D may be a preferred indicator of vitamin D status compared with total serum vitamin D, especially in conditions like pregnancy, liver disease or kidney disease that affect vitamin D binding protein levels.^{66 67} Given that the GWAS of vitamin D published to date used total serum vitamin D as the phenotype, we could not explore the effect of free vitamin D on COVID-19 in this study, and it is possible that genetic variation in free serum vitamin D is not captured by our genetic instruments. This limitation is mitigated given that free and total serum vitamin D are closely correlated in general and multiple clinical populations⁶⁸ and that there are limited data to support the claim that free or bioactive serum vitamin D is superior to total serum vitamin D as a predictor of vitamin D function, including effects on parathyroid hormone levels, calcium levels and bone health and metabolism.⁶⁹ Future research to investigate the differences between free and total serum vitamin D in relation to immune function is needed.

This study was not designed to evaluate the effect of acute changes in vitamin D status (ie, from supplementation) on prevention or treatment of COVID-19. One randomised trial and one quasi-experimental trial have provided evidence for a positive effect of vitamin D supplementation around time of diagnosis on COVID-19 prognosis in infected individuals. The doses of vitamin D administered in these studies could result in an acute change in the availability of vitamin D, which may support the immune system's response to the virus, mitigate acute lung injury and contribute to improved prognosis. Our results, which pertain to long-term vitamin D status, do not preclude the possibility that therapeutic doses of vitamin D may be effective in preventing or treating COVID-19 infection. Larger randomised trials using diverse sample populations are needed to investigate the potential use of therapeutic doses of vitamin D supplementation for COVID-19 prevention and treatment.

In conclusion, we used two-sample MR to study the associations of vitamin D with COVID-19 outcomes to address limitations of existing observational studies, including confounding and reverse causality. We found no evidence for an effect of genetically predicted variation in serum vitamin D on risk or severity of COVID-19 infection in a largely European population. Our findings suggest that

What this paper adds

What is already known on this topic

- ▶ Observational studies report that lower vitamin D levels are associated with increased risk and severity of COVID-19.
- ▶ Known risk factors for COVID-19 are associated with lower vitamin D levels in blood.
- ▶ It is unknown whether observed associations of vitamin D and COVID-19 are causal.

What this study adds

- ▶ This study found no evidence for associations of genetically predicted long-term vitamin D levels with risk and severity of COVID-19 infection in a largely European population.
- ▶ The findings do not support a causal effect of long-term usual vitamin D nutritional status on COVID-19 infection and its severity.
- ▶ These findings have implications for the potential effectiveness of low-level vitamin D supplementation as a strategy for COVID-19 prevention and treatment, but do not address the therapeutic use of vitamin D in acute disease.

chronic differences in serum vitamin D do not have a causal effect on susceptibility to COVID-19 infection or severity of COVID-19 among those infected, and that associations observed in previous studies may have been driven by confounding. Future directions of this work include extension of the MR approach to non-European ancestry populations, and investigation of potential modification of genetically predicted vitamin D effects on COVID-19 risk and severity by COVID-19 risk factors, including race/ethnicity, age, sex and BMI. Randomised trials are needed to inform the question of whether acute changes in vitamin D levels (ie, from supplementation) are efficacious in COVID-19 prevention and treatment across diverse populations.

Contributors The study was designed and conceived by BP, PC and AGC. The MR analyses were performed by BP in close consultation with PC. The follow-up stratified analyses of SNP-vitamin D associations were performed by NG. The manuscript was drafted by BP and PC, and edited with input from AGC, DBH and NG. All authors reviewed and contributed to the discussion of findings and the crafting of the manuscript and gave final approval to the version submitted for publication.

Funding BP was supported by the National Institutes of Health under award T32-DK007158. AGC was supported by the National Institutes of Health under award R01-HG006849. NG, DBH and PC were partially supported by the National Institutes of Health under award R01-HL149352.

Disclaimer The study content is solely the responsibility of the authors and does not necessarily represent the official views of the funders. The funders had no role in the study design, data collection, analysis, interpretation, or writing, nor in the decision to submit the article for publication.

Competing interests All authors have completed the ICMJE uniform disclosure form at www.icmje.org/doi_disclosure.pdf and declare no support from any organisation for the submitted work. BP was supported by the National Institutes of Health during the conduct of the study; AGC, DBH and PC report receiving grants from the National Institutes of Health during the conduct of the study; no other relationships or activities that could appear to have influenced the submitted work.

Patient consent for publication Not required.

Ethics approval The main analysis conducted in this study used publicly available summary data and did not require ethical approval. The use of UK Biobank data for the stratified analysis of SNP-vitamin D associations by COVID-19 risk factors was approved by the Institutional Review Board at RTI International.

Provenance and peer review Not commissioned; externally peer reviewed by Xia Jiang, Karolinska Institute, Sweden and Dr Emmanuel Baah, University of North Carolina System.

Data availability statement All data used for this analysis are publicly available. UK Biobank and COVID-19 Host Genetics Initiative data used for the main analysis are available in a public, open access repository. UK Biobank data used for stratified analyses of SNP-vitamin D associations are available on reasonable request from the UK Biobank. Code implementing the MR analysis is available on request from the corresponding author.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Bonnie K Patchen <http://orcid.org/0000-0003-3361-6744>

Patricia A Cassano <http://orcid.org/0000-0003-4827-5073>

REFERENCES

- 1 Home. Johns hopkins coronavirus resource center. Available: <https://coronavirus.jhu.edu/>
- 2 Guillot X, Semerano L, Saidenberg-Kermanac'h N, et al. Vitamin D and inflammation. *Joint Bone Spine* 2010;77:552–7.
- 3 Brenner H, Holleczer B, Schöttker B. Vitamin D insufficiency and deficiency and mortality from respiratory diseases in a cohort of older adults: potential for limiting the death toll during and beyond the COVID-19 pandemic? *Nutrients* 2020;12:2488.
- 4 Cui C, Xu P, Li G, et al. Vitamin D receptor activation regulates microglia polarization and oxidative stress in spontaneously hypertensive rats and angiotensin II-exposed microglial cells: role of renin-angiotensin system. *Redox Biol* 2019;26:101295.
- 5 Kai H, Kai M. Interactions of coronaviruses with ACE2, angiotensin II, and Ras inhibitors—lessons from available evidence and insights into COVID-19. *Hypertens Res* 2020;1–7.
- 6 Mok CK, Ng YL, Ahidjo BA. Calcitriol, the active form of vitamin D, is a promising candidate for COVID-19 prophylaxis. *bioRxiv*.
- 7 Entrenas Castillo M, Entrenas Costa LM, Vaquero Barrios JM, et al. "Effect of calcifediol treatment and best available therapy versus best available therapy on intensive care unit admission and mortality among patients hospitalized for COVID-19: A pilot randomized clinical study". *J Steroid Biochem Mol Biol* 2020;203:105751.
- 8 Annweiler C, Hanotte B, Grandin de l'Epreux C, et al. Vitamin D and survival in COVID-19 patients: a quasi-experimental study. *J Steroid Biochem Mol Biol* 2020;204:105771.
- 9 Ilie PC, Stefanescu S, Smith L. The role of vitamin D in the prevention of coronavirus disease 2019 infection and mortality. *Aging Clin Exp Res* 2020;32:1195–8.
- 10 Whittemore PB. COVID-19 fatalities, latitude, sunlight, and vitamin D. *Am J Infect Control* 2020;48:1042–4.
- 11 Kohlmeier M. Avoidance of vitamin D deficiency to slow the COVID-19 pandemic. *BMJ Nutr Prev Health* 2020;3:bmjnp-2020-000096.
- 12 Kaufman HW, Niles JK, Kroll MH, et al. SARS-CoV-2 positivity rates associated with circulating 25-hydroxyvitamin D levels. *PLoS One* 2020;15:e0239252.
- 13 Meltzer DO, Best TJ, Zhang H, et al. Association of vitamin D status and other clinical characteristics with COVID-19 test results. *JAMA Netw Open* 2020;3:e2019722.
- 14 D'Avolio A, Avataneo V, Manca A, et al. 25-Hydroxyvitamin D concentrations are lower in patients with positive PCR for SARS-CoV-2. *Nutrients* 2020;12:1359.
- 15 Ye K, Tang F, Liao X, et al. Does serum vitamin D level affect COVID-19 infection and its severity?—a case-control study. *J Am Coll Nutr* 2020;1–8.
- 16 Maghbooli Z, Sahraian MA, Ebrahimi M, et al. Vitamin D sufficiency, a serum 25-hydroxyvitamin D at least 30 ng/mL reduced risk for adverse clinical outcomes in patients with COVID-19 infection. *PLoS One* 2020;15:e0239799.
- 17 Macaya F, Espejo Paeres C, Valls A, et al. Interaction between age and vitamin D deficiency in severe COVID-19 infection. *Nutr Hosp* 2020;37:1039–42.
- 18 Baktash V, Hosack T, Patel N, et al. Vitamin D status and outcomes for hospitalised older patients with COVID-19. *Postgrad Med J* 2020. doi:10.1136/postgradmedj-2020-138712. [Epub ahead of print: 27 Aug 2020].
- 19 COVID-19 Treatment Guidelines Panel. Coronavirus disease 2019 (COVID-19) treatment guidelines. COVID-19 treatment guidelines. Available: <https://www.covid19treatmentguidelines.nih.gov/adjunctive-therapy/vitamin-d/>
- 20 Williams C, Gingery JG. Joint guidance on vitamin D in the era of COVID-19. Endocrine Society, 2020. Available: <https://www.endocrine.org/news-and-advocacy/news-room/2020/joint-guidance-on-vitamin-d>
- 21 Hastie CE, Mackay DF, Ho F, et al. Vitamin D concentrations and COVID-19 infection in UK Biobank. *Diabetes Metab Syndr* 2020;14:561–5.
- 22 Manousaki D, Mitchell R, Dudding T, et al. Genome-wide association study for vitamin D levels reveals 69 independent loci. *Am J Hum Genet* 2020;106:327–37.
- 23 Jiang X, O'Reilly PF, Aschard H, et al. Genome-wide association study in 79,366 European-ancestry individuals informs the genetic architecture of 25-hydroxyvitamin D levels. *Nat Commun* 2018;9:260.
- 24 Wang TJ, Zhang F, Richards JB, et al. Common genetic determinants of vitamin D insufficiency: a genome-wide association study. *Lancet* 2010;376:180–8.
- 25 Ahn J, Yu K, Stolzenberg-Solomon R, et al. Genome-wide association study of circulating vitamin D levels. *Hum Mol Genet* 2010;19:2739–45.
- 26 Hong J, Hatchell KE, Bradfield JP, et al. Transethnic evaluation identifies low-frequency loci associated with 25-hydroxyvitamin D concentrations. *J Clin Endocrinol Metab* 2018;103:1380–92.
- 27 Arabi A, Khoeiry-Zgheib N, Awada Z, et al. CYP2R1 polymorphisms are important modulators of circulating 25-hydroxyvitamin D levels in elderly females with vitamin insufficiency, but not of the response to vitamin D supplementation. *Osteoporos Int* 2017;28:279–90.
- 28 Jones P, Luccock M, Chaplin G, et al. Distribution of variants in multiple vitamin D-related loci (DHCR7/NADSYN1, GC, CYP2R1, CYP11A1, CYP24A1, VDR, RXRα and RXRγ) vary between European, East-Asian and sub-Saharan African-ancestry populations. *Genes Nutr* 2020;15:5.
- 29 Liefwaard MC, Ligthart S, Vitezova A, et al. Vitamin D and C-reactive protein: a Mendelian randomization study. *PLoS One* 2015;10:e0131740.
- 30 Afzal S, Brøndum-Jacobsen P, Bojesen SE, et al. Genetically low vitamin D concentrations and increased mortality: Mendelian randomisation analysis in three large cohorts. *BMJ* 2014;349:g6330.
- 31 Brøndum-Jacobsen P, Benn M, Afzal S, et al. No evidence that genetically reduced 25-hydroxyvitamin D is associated with increased risk of ischaemic heart disease or myocardial infarction: a Mendelian randomization study. *Int J Epidemiol* 2015;44:651–61.
- 32 Bae S-C, Lee YH. Vitamin D level and risk of systemic lupus erythematosus and rheumatoid arthritis: a Mendelian randomization. *Clin Rheumatol* 2018;37:2415–21.
- 33 Hysinger EB, Roizen JD, Mentch FD, et al. Mendelian randomization analysis demonstrates that low vitamin D is unlikely causative for pediatric asthma. *J Allergy Clin Immunol* 2016;138:1747–9.
- 34 Manousaki D, Mokry LE, Ross S, et al. Mendelian randomization studies do not support a role for vitamin D in coronary artery disease. *Circ Cardiovasc Genet* 2016;9:349–56.
- 35 Manousaki D, Paternoster L, Standl M, et al. Vitamin D levels and susceptibility to asthma, elevated immunoglobulin E levels, and atopic dermatitis: a Mendelian randomization study. *PLoS Med* 2017;14:e1002294.
- 36 Mao Y, Zhan Y, Huang Y. Vitamin D and asthma: a Mendelian randomization study. *Ann Allergy Asthma Immunol* 2017;119:95–7.
- 37 Mokry LE, Ross S, Morris JA, et al. Genetically decreased vitamin D and risk of Alzheimer disease. *Neurology* 2016;87:2567–74.
- 38 Ong J-S, Cuellar-Partida G, Lu Y, et al. Association of vitamin D levels and risk of ovarian cancer: a Mendelian randomization study. *Int J Epidemiol* 2016;45:1619–30.
- 39 Skaaby T, Husemoen LLN, Martinussen T, et al. Vitamin D status, filaggrin genotype, and cardiovascular risk factors: a Mendelian randomization approach. *PLoS One* 2013;8:e57647.

- 40 Taylor AE, Burgess S, Ware JJ, *et al.* Investigating causality in the association between 25(OH)D and schizophrenia. *Sci Rep* 2016;6:26496.
- 41 Trummer O, Pilz S, Hoffmann MM, *et al.* Vitamin D and mortality: a Mendelian randomization study. *Clin Chem* 2013;59:793–7.
- 42 Viatte S, Yarwood A, McAllister K, *et al.* The role of genetic polymorphisms regulating vitamin D levels in rheumatoid arthritis outcome: a Mendelian randomisation approach. *Ann Rheum Dis* 2014;73:1430–3.
- 43 Vimalaewaran KS, Berry DJ, Lu C, *et al.* Causal relationship between obesity and vitamin D status: bi-directional Mendelian randomization analysis of multiple cohorts. *PLoS Med* 2013;10:e1001383.
- 44 Vimalaewaran KS, Cavadino A, Berry DJ, *et al.* Association of vitamin D status with arterial blood pressure and hypertension risk: a Mendelian randomisation study. *Lancet Diabetes Endocrinol* 2014;2:719–29.
- 45 Ye Z, Sharp SJ, Burgess S, *et al.* Association between circulating 25-hydroxyvitamin D and incident type 2 diabetes: a Mendelian randomisation study. *Lancet Diabetes Endocrinol* 2015;3:35–42.
- 46 Amin HA, Drenos F. No evidence that vitamin D is able to prevent or affect the severity of COVID-19 in individuals with European ancestry: a Mendelian randomisation study of open data. *BMJ Nutr Prev Health* 2021;8:bmjnp-2020-000151.
- 47 Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013;37:658–65.
- 48 COVID-19 Host Genetics Initiative. The COVID-19 host genetics initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur J Hum Genet* 2020;28:1–4.
- 49 Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol* 2016;40:597–608.
- 50 Burgess S, Thompson SG, CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol* 2011;40:755–64.
- 51 Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 2015;31:3555–7.
- 52 Hoffmann TJ, Choquet H, Yin J, *et al.* A large multiethnic genome-wide association study of adult body mass index identifies novel loci. *Genetics* 2018;210:499–515.
- 53 Shungin D, Winkler TW, Croteau-Chonka DC, *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 2015;518:187–96.
- 54 Liu M, Jiang Y, Wedow R, *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* 2019;51:237–44.
- 55 Locke AE, Kahali B, Berndt SI, *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015;518:197–206.
- 56 Hemani G, Zheng J, Elsworth B, *et al.* The MR-Base platform supports systematic causal inference across the human phenotype. *Elife* 2018;7:e34408.
- 57 Gogarten SM, Sofer T, Chen H, *et al.* Genetic association testing using the genesis R/Bioconductor package. *Bioinformatics* 2019;35:5346–8.
- 58 Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000;155:945–59.
- 59 1000 Genomes Project Consortium, Auton A, Brooks LD, *et al.* A global reference for human genetic variation. *Nature* 2015;526:68–74.
- 60 Nature. An integrated map of structural variation in 2,504 human genomes. Available: <https://www.nature.com/articles/nature15394>
- 61 Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet* 2018;27:R195–208.
- 62 Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;44:512–25.
- 63 Bowden J, Davey Smith G, Haycock PC, *et al.* Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016;40:304–14.
- 64 Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* 2017;46:1985–98.
- 65 Brion M-JA, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol* 2013;42:1497–501.
- 66 Bikle DD, Schwartz J. Vitamin D binding protein, total and free vitamin D levels in different physiological and pathophysiological conditions. *Front Endocrinol* 2019;10:317.
- 67 Bikle DD. The free hormone hypothesis: when, why, and how to measure the free hormone levels to assess vitamin D, thyroid, sex hormone, and cortisol status. *JBMR Plus* 2021;5:e10418.
- 68 Schwartz JB, Gallagher JC, Jorde R, *et al.* Determination of free 25(OH)D concentrations and their relationships to total 25(OH)D in multiple clinical populations. *J Clin Endocrinol Metab* 2018;103:3278–88.
- 69 Tsuprykov O, Chen X, Hoche C-F, *et al.* Why should we measure free 25(OH) vitamin D? *J Steroid Biochem Mol Biol* 2018;180:87–104.

Supplemental Table 1. Characteristics of SNPs included in Instrument C

Gene	SNP	EA	NEA	EAF	BETA	P-VAL	F STAT	Non-vitamin D GWAS associations
GC	rs11723621	G	A	0.29	-0.19	2.9E-1689	1467	YES
CYP2R1	rs10832289	T	A	0.41	-0.07	2.03E-266	231	YES
NADSYN/DHCR7	rs12803256	G	A	0.78	0.10	1.3E-378	354	YES
CYP24A1	rs6127099	T	A	0.28	-0.04	9.30E-62	55	YES
SEC23A	rs8018720	C	G	0.82	-0.03	4.04E-36	30	NO
AMDHD1/HAL	rs10859995	C	T	0.58	-0.04	7.03E-89	76	YES
RER1	rs6698680	G	A	0.46	-0.01	8.99E-10	7	YES
PADI1	rs3750296	C	G	0.34	-0.02	2.09E-24	19	NO
RP4-657M3.2	rs7519574	A	G	0.18	0.02	2.09E-11	9	NO
FOXO6	rs56044892	T	C	0.21	0.02	2.85E-10	8	YES
DOCK7	rs2934744	A	C	0.64	-0.02	3.96E-26	23	YES
CELSR2	rs7528419	G	A	0.22	0.02	2.41E-16	13	YES
ARNT	rs3768013	A	G	0.37	-0.01	1.37E-13	10	YES
FLG	rs61816761	A	G	0.02	0.13	8.57E-74	71	YES
FDPS	rs11264360	A	T	0.24	0.02	3.34E-15	12	YES
MARC_1	rs867772	G	A	0.68	-0.01	3.64E-11	8	YES
-NA-	rs10127775	T	A	0.60	0.01	3.43E-09	7	YES
TDRD15	rs12997242	A	G	0.44	-0.01	2.23E-10	8	YES
GCKR	rs11127048	A	G	0.62	0.02	6.41E-19	16	YES
NPAS2	rs6724965	G	A	0.17	-0.02	1.29E-10	8	NO
HTR5BP	rs7569755	A	G	0.29	0.01	8.03E-11	8	NO
CPS1	rs1047891	A	C	0.32	-0.01	1.16E-11	9	YES
UGT1A4	rs2011425	G	T	0.08	-0.05	9.66E-38	31	YES
RHOA	rs7650253	A	T	0.69	0.01	1.76E-10	9	YES
CADM2	rs1972994	T	A	0.65	-0.02	7.99E-18	14	YES
MRPL3	rs6438900	G	C	0.26	0.01	9.59E-10	7	YES
TFDP2	rs6773343	T	C	0.72	0.01	5.20E-09	6	YES
DOK7	rs78649910	A	T	0.11	-0.02	4.32E-09	7	YES
UGT2B7	rs7699711	T	G	0.45	-0.03	6.97E-49	41	YES
HSD17B11	rs58073039	G	A	0.30	-0.01	2.16E-11	8	NO
ADH1A	rs28364331	G	A	0.02	0.06	1.31E-17	14	YES
TNFAIP8	rs7718395	G	C	0.32	0.01	1.67E-09	7	NO
MED23	rs3822868	G	A	0.84	0.02	1.41E-15	13	YES
DNAH11	rs111529171	C	G	0.22	-0.02	6.24E-11	8	YES
LINC01004	rs1011468	A	G	0.48	-0.01	1.35E-12	9	YES

Gene	SNP	EA	NEA	EAF	BETA	P-VAL	F STAT	Non-vitamin D GWAS associations
COG5	rs1858889	C	A	0.50	0.01	3.85E-11	8	YES
GATA4	rs804280	A	C	0.58	0.01	4.43E-11	8	YES
EBF2	rs34726834	T	C	0.25	0.01	6.65E-10	7	YES
LINC00536	rs7828742	G	A	0.60	-0.02	3.06E-28	23	YES
DNAH11	rs10818769	G	C	0.86	-0.02	3.35E-09	7	YES
ABO	rs532436	A	G	0.18	-0.02	2.17E-09	7	YES
MAT1A	rs10887718	T	C	0.53	-0.01	1.44E-10	8	YES
PLEKHA7	rs567415847	G	A	1.00	0.28	1.03E-14	32	YES
TMEM151A	rs523583	C	A	0.47	0.01	5.58E-10	7	YES
RP11-21L23.4	rs1149605	C	T	0.17	0.02	7.34E-14	11	YES
ZPR1	rs964184	C	G	0.86	0.04	5.11E-44	37	YES
SLCO1B1	rs12317268	G	A	0.15	-0.02	9.15E-12	9	YES
FAM166AP9	rs9668081	T	C	0.47	0.01	5.38E-09	7	YES
LIPC	rs1800588	T	C	0.21	-0.03	2.65E-36	30	YES
AC007950.2	rs17765311	C	A	0.34	-0.02	1.35E-13	10	YES
PEAK1	rs62007299	A	G	0.71	-0.01	1.69E-11	9	YES
BCAR4	rs8063706	T	A	0.27	0.01	3.64E-09	7	NO
PDILT	rs77924615	A	G	0.20	-0.02	1.46E-10	8	YES
FBXL19	rs71383766	T	C	0.42	0.01	1.15E-09	8	YES
CETP	rs1800775	A	C	0.49	-0.02	1.56E-17	14	YES
RP11-120M18.2	rs2909218	T	C	0.79	0.02	2.81E-12	9	YES
DSG1	rs8091117	A	C	0.07	-0.02	1.03E-09	7	NO
SERPINB11	rs2037511	A	G	0.17	0.02	9.29E-10	7	YES
STAP2	rs57631352	G	A	0.30	-0.01	1.48E-09	7	YES
LDLR	rs73015021	G	A	0.12	0.02	1.15E-14	11	YES
TM6SF2	rs58542926	T	C	0.08	0.03	8.57E-19	15	YES
NPHS1	rs3814995	T	C	0.31	-0.01	2.83E-12	9	NO
APOC1	rs157595	G	A	0.61	-0.02	2.95E-14	12	YES
SULT2A1	rs112285002	T	C	0.16	0.06	1.77E-110	98	YES
KLK10	rs10426	A	G	0.21	0.03	3.31E-26	21	YES
ZNF808	rs8103262	C	T	0.31	0.01	3.18E-09	7	NO
NRIP1	rs2229742	C	G	0.10	-0.03	7.13E-16	12	YES
PLA2G3	rs2074735	C	G	0.06	0.03	6.55E-12	9	YES
SCUBE1	rs960596	T	C	0.34	0.01	2.23E-09	7	YES

BETA=beta coefficient for association with vitamin D in UKBB; P-VAL=p-value for SNP—vitamin D association in UKBB; F-stat= measure of instrument strength, calculated as described in the methods, shown for sample size of 100,000; Non-vitamin D GWAS associations = indicator of whether SNP has known genome-wide significant associations with traits other than vitamin D, identified as described in methods using LDlink webtool

Supplemental Table 2) P-values for associations of vitamin D SNPs with BMI and smoking phenotypes

Vitamin D SNP	Locus	BMI (continuous) ^a	Waist Circumference ^a	Obesity (case/control) ^a	Smoking (ever vs never) ^b	Smoking (former vs current) ^b	Smoking (cigarettes per day) ^b
rs11723621	GC	0.500	0.720 ^c	0.6376	0.869	0.645	0.751
rs10832289	CYP2R1	0.00017	0.0023	0.3962	0.462	0.701	0.165
rs12803256	DHCR7	0.590	0.470	0.9025	0.152	0.369	0.353
rs6127099	CYP24A1	0.490	0.440 ^c	0.5857	0.707	0.302	0.517
rs8018720	SEC23A	0.390	0.680	0.2976	0.294	0.133	0.0123
rs10859995	AMDHD1	0.830	0.760	0.2801	0.169	0.17	0.517

^a P-values for associations with BMI, waist circumference, and obesity accessed through MR-Base platform using the following study IDs: “ebi-a-GCST006368” (BMI) ; “ieu-a-60” (waist circumference); “finn-a-E4_OBESITY” (obesity).

^b P-values for associations with smoking phenotypes from Liu et al, 2019, (<https://www.nature.com/articles/s41588-018-0307-5>).

^c Data were not available for rs11723621 and rs6127099 for the waist circumference outcome. P-values shown are for the GC and CYP24A1 SNPs identified in the SUNLIGHT Consortium (rs3755967 and rs17216707)

Supplemental Table 3. Studies contributing to different COVID-19 outcomes in the COVID-19 Host Genetics Initiative

Study	Percent of COVID-19 outcome sample				
	COVID vs. Population	Hosp COVID vs. Population	Sev Resp COVID vs. Population	COVID Pos vs. COVID Neg	Hosp vs. non-hosp COVID
Amsterdam UMC COVID study group	0.11%	0.16%	0.24%		
Ancestry	1.25%	0.23%		13.57%	20.32%
Genetic modifiers for COVID-19 related illness (BelCovid)	0.11%	0.16%			
BoSCO		0.04%	0.05%		3.68%
Biobanque Quebec COVID19 (BQC19)	0.04%	0.06%	0.09%	0.42%	1.89%
Genetic determinants of COVID-19 complications in the Brazilian population	0.17%	0.25%	0.33%		
Genetics of COVID-related Manifestation (Corea)	0.48%	0.68%			
deCODE	19.82%	28.30%		24.17%	17.39%
Estonian Biobank	9.98%			9.64%	
FinnGen	17.19%	24.62%	38.00%		2.81%
GEN-COVID	0.23%	0.31%	0.47%		
genomiCC	0.72%	1.04%	1.60%		
Genomics England (genomicsengland100kcp)	4.50%			1.43%	
Genes & Health (GNH)	1.97%	2.83%		0.29%	0.94%
Helix Exome+ COVID-19 Phenotypes	0.40%				
COVID19-Host(a)ge	0.27%	0.39%			
UK Blood Donors Cohort	3.01%			1.00%	
Italy COVID19-Host(a)ge			0.31%		
Lifelines	1.84%			1.26%	
Michigan Genomics Initiative	3.71%			0.49%	

Study	Percent of COVID-19 outcome sample				
	COVID vs. Population	Hosp COVID vs. Population	Sev Resp COVID vs. Population	COVID Pos vs. COVID Neg	Hosp vs. non-hosp COVID
Million Veterans Program	1.40%	0.56%		32.96%	29.69%
Netherlands Twin Register	0.39%			0.20%	
Partners Healthcare Biobank	2.53%			3.16%	
Penn Medicine Biobank	0.62%	0.89%		0.86%	1.52%
Qatar Genome Program	1.01%	1.38%			6.42%
Spain COVID19-Host(a)ge			0.20%		
Determining the Molecular Pathways and Genetic Predisposition of the Acute Inflammatory Process Caused by SARS-CoV-2	0.05%	0.06%	0.06%		3.32%
Genomic epidemiology of SARS-Cov-2 and host genetics in Coronavirus Disease 2019 (COVID-19)	0.02%				
The genetic predisposition to severe COVID-19	0.28%	0.40%	0.61%		
UK Biobank	27.87%	37.65%	58.05%	12.28%	12.01%

Supplemental Table 4. Minimum detectable odds ratios for COVID-19 outcomes

COVID-19 outcome	N cases	N controls	Case prevalence	Min detectable Odds Ratio*
COVID vs population	17,965	1,370,547	0.013	0.88
Hosp COVID vs pop	7,885	961,804	0.008	0.82
Severe resp COVID vs pop	4,336	623,902	0.007	0.73
COVID pos vs COVID neg	110,85	116,794	0.087	0.84
Hosp COVID vs non-hosp COVID	2,430	8,478	0.223	0.66

*Minimum detectable odds ratios for 80% power, represents odds ratio per SD increase in serum vitamin D, assumes 3% variance in vitamin D explained by genetic instruments

Supplemental Table 5. Multivariable MR estimates for direct effect of vitamin D on COVID-19 outcome conditioning on BMI

MR method	Conditioning GWAS		COVID vs. population		Hosp COVID vs. population		Severe Resp COVID vs. population		COVID pos vs. COVID neg		Hosp COVID vs. non-hosp COVID	
	Study ID	Ancestry	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
MVMR	ebi-a-GCST006368	European	1.04	0.93-1.17	1.14	0.96-1.35	1.01	0.79-1.31	1.13	0.97-1.32	1.36	0.90-2.06
MVMR	ieu-a-835	European	1.04	0.91-1.18	1.15	0.93-1.43	0.99	0.74-1.33	1.14	0.97-1.33	1.46	0.92-2.30
MVMR	ieu-a-2	Mixed	1.04	0.92-1.16	1.15	0.95-1.39	1.00	0.76-1.30	1.14	0.96-1.34	1.48	1.01-2.16

OR, odds ratios per unit increase in serum vitamin D in natural log scale; 95% CI, 95% confidence interval

Supplemental Table 6. MR analysis limiting instrument to SNPs with no known associations except with vitamin D

MR method	COVID vs. population		Hosp COVID vs. population		Severe Resp COVID vs. population		COVID pos vs. COVID neg		Hosp COVID vs. non-hosp COVID	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
IVW	0.94	0.55-1.62	0.92	0.34-2.50	0.92	0.27-3.15	0.96	0.51-1.80	0.60	0.12-2.98
MR-Egger intercept test	p-value = 0.32		p-value = 0.62		p-value = 0.89		p-value = 0.17		p-value = 0.92	

OR, odds ratios per SD increase in log-transformed serum vitamin D; 95% CI, 95% confidence interval;

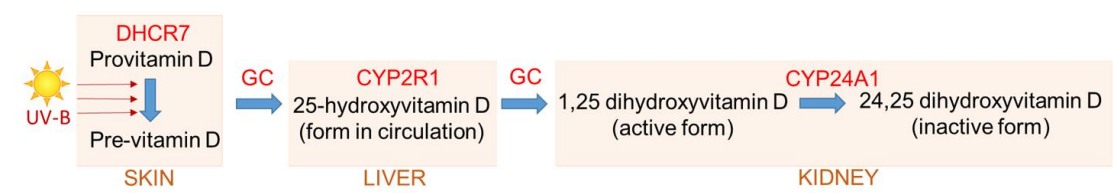
SNPs included in instrument: rs3750296, rs3814995, rs58073039, rs6724965, rs7519574, rs8103262, rs7569755, rs7718395, rs8018720, rs8063706, rs8091117

Supplemental Table 7. Replication of MR estimates of effect of vitamin D on COVID-19 outcomes using SNP-vitamin D association data from SUNLIGHT Consortia

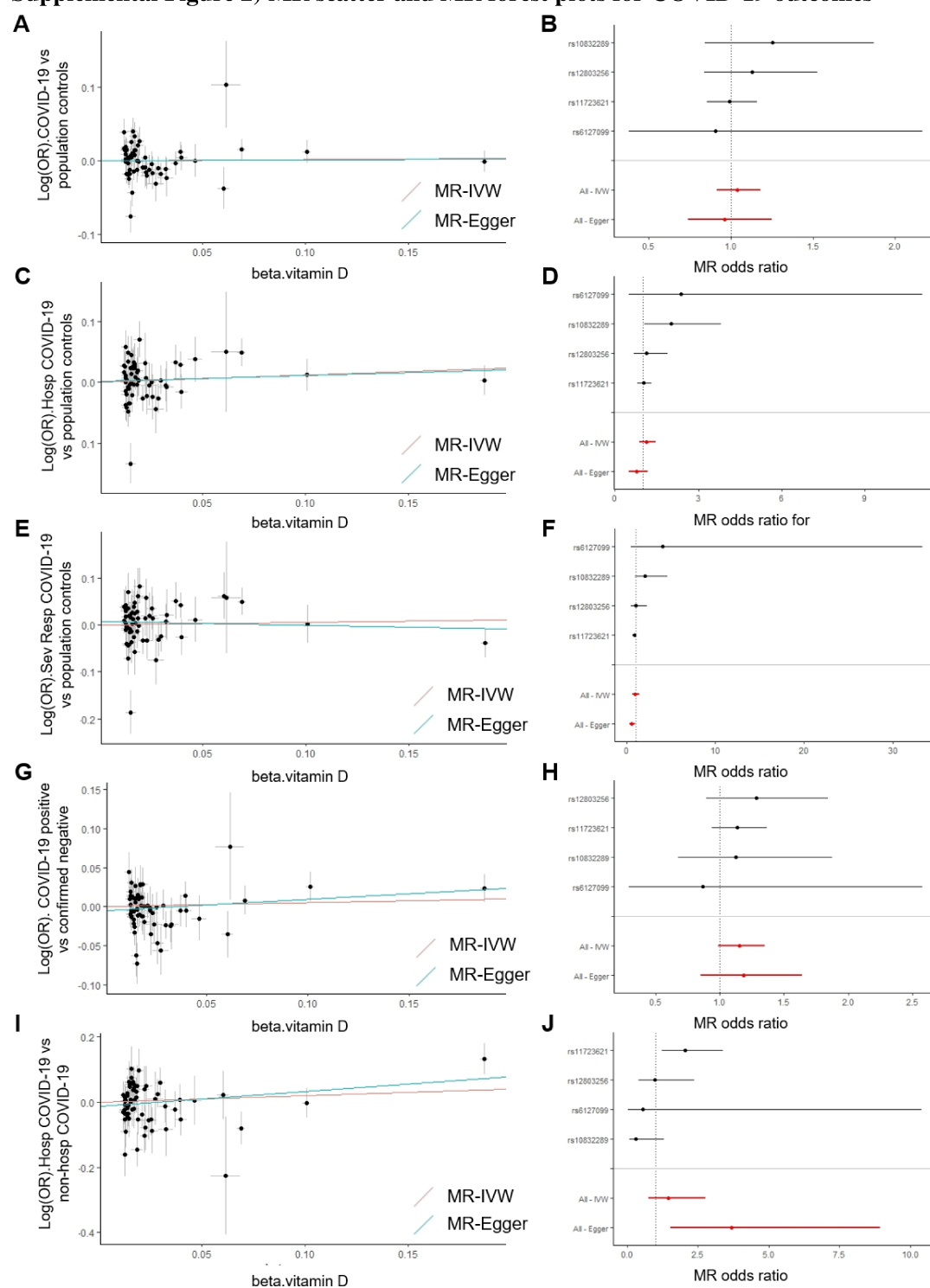
MR method	COVID vs. population		Hosp COVID vs. population		Severe Resp COVID vs. population		COVID pos vs. COVID neg		Hosp COVID vs. non-hosp COVID	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Instrument A										
IVW	1.04	0.79-1.35	1.04	0.67-1.62	0.72	0.40-1.29	1.16	0.77-1.75	2.39	0.93-6.13
MR-Egger intercept test	p-value = 0.90		p-value = 0.91		p-value = 0.40		p-value = 0.26		p-value = 0.13	

OR, odds ratios per unit increase in serum vitamin D in natural log scale; 95% CI, 95% confidence interval

Supplemental Figure 1) Genetic Loci integral to the vitamin D metabolic pathway

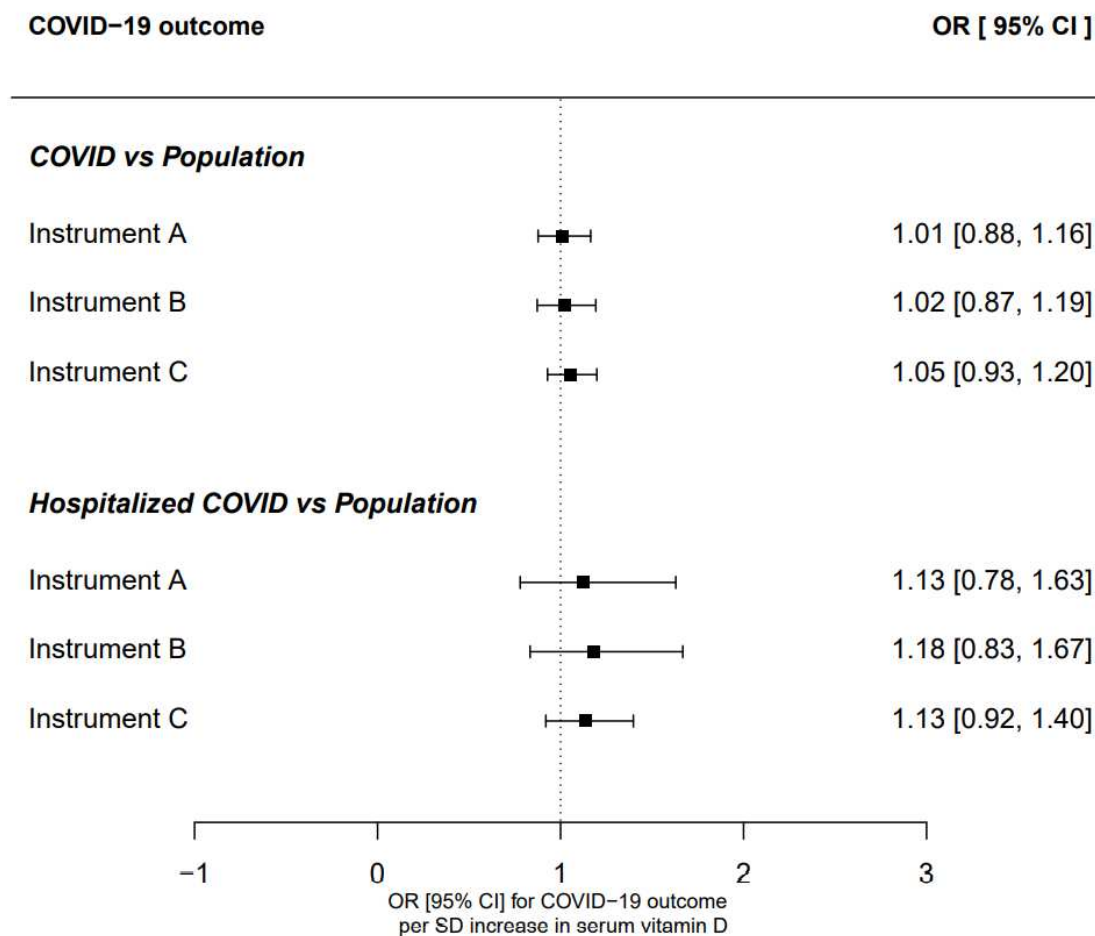


Supplemental Figure 2) MR scatter and MR forest plots for COVID-19 outcomes

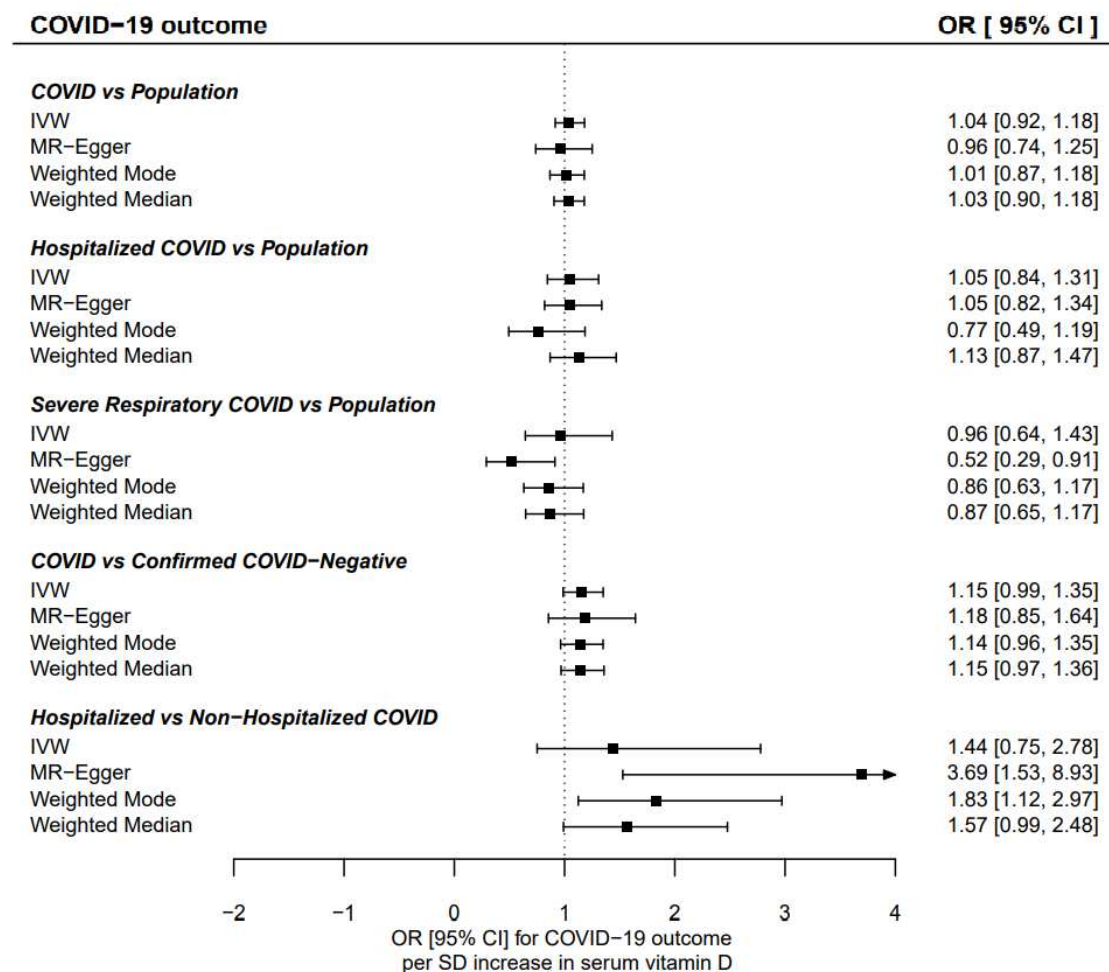


Supplemental Figure 2 caption: MR scatter and forest plots for (A-B) COVID-19 vs population controls, (C-D) Hospitalized COVID-19 vs population controls, (E-F) Severe respiratory COVID-19 vs population controls, (G-H) COVID-19 vs confirmed negative and (I-J) hospitalized COVID-19 vs non-hospitalized COVID-19. Scatter plots show the SNP—vitamin D associations on the x-axis vs the SNP—log(OR) for COVID-19 case vs comparator on the y axis for all SNPs considered in instruments A,B and C. The MR estimates generated from the IVW (red slope) and MR-Egger (teal slope) MR methods are overlaid. The forest plots show the MR estimates (odds ratios and 95% confidence intervals) generated from instrument A for the effect of 1 SD higher serum vitamin D on risk of COVID-19 case vs comparator. Estimates for each SNP are shown separately (black bars), as well as summarized estimates across all SNPs (red bars).

Supplemental Figure 3) Inverse variance weighted MR estimates of effect of vitamin D on risk of COVID-19 infection and hospitalization in European ancestry participants



Supplemental Figure 4) MR estimates of effect of vitamin D on COVID-19 outcomes using different MR Methods for instrument A



Supplemental Figure 5) Pairwise correlations of beta-coefficients observed for SNP-vitamin D associations stratified by COVID-19 risk subgroup, across all SNPs in instruments A, B and C

